

## RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.

Application Serial Number: 10/524,053A  
Source: PU  
Date Processed by STIC: 1/27/06

# ***ENTERED***



PCT

## RAW SEQUENCE LISTING

DATE: 01/27/2006

PATENT APPLICATION: US/10/524,053A

TIME: 14:44:07

Input Set : A:\DOCSMTL-#1912723-v1-SEQLIST2.TXT

Output Set: N:\CRF4\01272006\J524053A.raw

4 <110> APPLICANT: NATIONAL RESEARCH COUNCIL OF CANADA  
 5 SU, Zhengding  
 6 NI, Feng  
 8 <120> TITLE OF INVENTION: Novel Fusion Proteins for the Production  
 9 of Recombinant Peptides  
 12 <130> FILE REFERENCE: 2139-22US  
 C--> 14 <140> CURRENT APPLICATION NUMBER: US/10/524,053A  
 C--> 14 <141> CURRENT FILING DATE: 2005-02-08  
 14 <150> PRIOR APPLICATION NUMBER: US 60/402,075  
 15 <151> PRIOR FILING DATE: 2002-08-09  
 17 <160> NUMBER OF SEQ ID NOS: 48  
 19 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
 21 <210> SEQ ID NO: 1  
 22 <211> LENGTH: 25  
 23 <212> TYPE: PRT  
 24 <213> ORGANISM: Staphylococcus nuclease  
 26 <400> SEQUENCE: 1  
 27 Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys  
 28 1 5 10 15  
 29 Ala Ile Asp Gly Asp Thr Val Lys Leu  
 30 20 25  
 32 <210> SEQ ID NO: 2  
 33 <211> LENGTH: 5  
 34 <212> TYPE: PRT  
 35 <213> ORGANISM: Staphylococcus nuclease  
 37 <400> SEQUENCE: 2  
 38 Tyr Lys Gly Gln Pro  
 39 1 5  
 41 <210> SEQ ID NO: 3  
 42 <211> LENGTH: 29  
 43 <212> TYPE: PRT  
 44 <213> ORGANISM: Staphylococcus nuclease  
 46 <400> SEQUENCE: 3  
 47 Leu Leu Leu Val Asp Thr Pro Glu Thr Lys His Pro Lys Lys Gly Val  
 48 1 5 10 15  
 49 Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe Thr Lys Lys  
 50 20 25  
 52 <210> SEQ ID NO: 4  
 53 <211> LENGTH: 32  
 54 <212> TYPE: PRT  
 55 <213> ORGANISM: Staphylococcus nuclease  
 57 <400> SEQUENCE: 4  
 58 Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp Lys Gly Gln Arg

## RAW SEQUENCE LISTING

DATE: 01/27/2006

PATENT APPLICATION: US/10/524,053A

TIME: 14:44:07

Input Set : A:\DOCSMTL-#1912723-v1-SEQLIST2.TXT

Output Set: N:\CRF4\01272006\J524053A.raw

59 1 5 10 15  
 60 Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr Ala Asp Gly Lys  
 61 20 25 30

63 &lt;210&gt; SEQ ID NO: 5

64 &lt;211&gt; LENGTH: 5

65 &lt;212&gt; TYPE: PRT

66 &lt;213&gt; ORGANISM: Staphylococcus nuclease

68 &lt;400&gt; SEQUENCE: 5

69 Val Asn Glu Ala Leu

70 1 5

72 &lt;210&gt; SEQ ID NO: 6

73 &lt;211&gt; LENGTH: 6

74 &lt;212&gt; TYPE: PRT

75 &lt;213&gt; ORGANISM: Staphylococcus nuclease

77 &lt;400&gt; SEQUENCE: 6

78 Phe Asn Pro Arg Gly Ser

79 1 5

81 &lt;210&gt; SEQ ID NO: 7

82 &lt;211&gt; LENGTH: 14

83 &lt;212&gt; TYPE: PRT

84 &lt;213&gt; ORGANISM: Staphylococcus nuclease

86 &lt;400&gt; SEQUENCE: 7

87 Val Arg Gln Gly Leu Ala Lys Val Ala Tyr Val Tyr Lys Pro

88 1 5 10

90 &lt;210&gt; SEQ ID NO: 8

91 &lt;211&gt; LENGTH: 32

92 &lt;212&gt; TYPE: PRT

93 &lt;213&gt; ORGANISM: Staphylococcus nuclease

95 &lt;400&gt; SEQUENCE: 8

96 Asn Asn Thr His Glu Gln Leu Leu Arg Lys Ser Glu Ala Gln Ala Lys

97 1 5 10 15

98 Lys Glu Lys Leu Asn Ile Trp Ser Glu Asp Asn Ala Asp Ser Gly Gln

99 20 25 30

102 &lt;210&gt; SEQ ID NO: 9

103 &lt;211&gt; LENGTH: 4

104 &lt;212&gt; TYPE: PRT

105 &lt;213&gt; ORGANISM: Artificial Sequence

107 &lt;220&gt; FEATURE:

108 <223> OTHER INFORMATION: Peptidic Cleavage Site found on the fusion carrier protein  
 derived from Staphylococcus nuclease

110 &lt;400&gt; SEQUENCE: 9

111 Phe Asn Pro Arg

112 1

114 &lt;210&gt; SEQ ID NO: 10

115 &lt;211&gt; LENGTH: 4

116 &lt;212&gt; TYPE: PRT

117 &lt;213&gt; ORGANISM: Artificial Sequence

119 &lt;220&gt; FEATURE:

120 <223> OTHER INFORMATION: Peptidic Cleavage Site found on the fusion carrier protein  
 derived from Staphylococcus nuclease

122 &lt;400&gt; SEQUENCE: 10

## RAW SEQUENCE LISTING

DATE: 01/27/2006

PATENT APPLICATION: US/10/524,053A

TIME: 14:44:07

Input Set : A:\DOCSMTL-#1912723-v1-SEQLIST2.TXT

Output Set: N:\CRF4\01272006\J524053A.raw

```

123 Leu Val Pro Arg
124 1
126 <210> SEQ ID NO: 11
128 <400> SEQUENCE: 11
W--> 129 000
131 <210> SEQ ID NO: 12
132 <211> LENGTH: 5
133 <212> TYPE: PRT
134 <213> ORGANISM: Artificial Sequence
136 <220> FEATURE:
137 <223> OTHER INFORMATION: Peptidic Cleavage Site found on the fusion carrier protein
derived from Staphylococcus nuclease
139 <400> SEQUENCE: 12
140 Asp Asp Asp Asp Lys
141 1 5
143 <210> SEQ ID NO: 13
144 <211> LENGTH: 351
145 <212> TYPE: DNA
146 <213> ORGANISM: Artificial Sequence
148 <220> FEATURE:
149 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120
151 <220> FEATURE:
152 <221> NAME/KEY: CDS
153 <222> LOCATION: (1)...(351)
155 <400> SEQUENCE: 13
156 gca act tca act aaa aaa tta cat aaa gaa cct gcg act tta att aaa 48
157 Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys
158 1 5 10 15
160 gcg att gat ggt gat acg gtt aaa tta atg tac aaa ggt caa cca atg 96
161 Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly Gln Pro Met
162 20 25 30
164 aca ttc aga cta tta ttg gtt gat aca cct gaa aca aag cat cct aaa 144
165 Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys His Pro Lys
166 35 40 45
168 aaa ggt gta gag aaa tat ggt cct gaa gca agt gca ttt acg aaa aaa 192
169 Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe Thr Lys Lys
170 50 55 60
172 atg gta gaa aat gca aag aaa att gaa gtc gag ttt gac aaa ggt caa 240
173 Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp Lys Gly Gln
174 65 70 75 80
176 aga act gat aaa tat gga cgt ggc tta gcg tat att tat gct gat gga 288
177 Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr Ala Asp Gly
178 85 90 95
180 aaa atg gta aac gaa gct tta gtt cgt caa ggc ttg gct aaa gtt gct 336
181 Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala Lys Val Ala
182 100 105 110
184 tat gtt tac aaa cct 351
185 Tyr Val Tyr Lys Pro
186 115
189 <210> SEQ ID NO: 14

```

## RAW SEQUENCE LISTING

DATE: 01/27/2006

PATENT APPLICATION: US/10/524,053A

TIME: 14:44:07

Input Set : A:\DOCSMTL-#1912723-v1-SEQLIST2.TXT

Output Set: N:\CRF4\01272006\J524053A.raw

```

190 <211> LENGTH: 117
191 <212> TYPE: PRT
192 <213> ORGANISM: Artificial Sequence
194 <220> FEATURE:
195 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120
197 <400> SEQUENCE: 14
198 Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys
199   1           5           10           15
200 Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly Gln Pro Met
201           20           25           30
202 Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys His Pro Lys
203           35           40           45
204 Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe Thr Lys Lys
205           50           55           60
206 Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp Lys Gly Gln
207 65           70           75           80
208 Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr Ala Asp Gly
209           85           90           95
210 Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala Lys Val Ala
211           100          105          110
212 Tyr Val Tyr Lys Pro
213           115
215 <210> SEQ ID NO: 15
216 <211> LENGTH: 369
217 <212> TYPE: DNA
218 <213> ORGANISM: Artificial Sequence
220 <220> FEATURE:
221 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120 ( N118E,
222   N119F)
224 <220> FEATURE:
226 <221> NAME/KEY: CDS
227 <222> LOCATION: (1)...(369)
229 <400> SEQUENCE: 15
230 aca gcc atg gca act tca act aaa aaa tta cat aaa gaa cct gcg act      48
231 Thr Ala Met Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr
232   1           5           10           15
234 tta att aaa gcg att gat ggt gat acg gtt aaa tta atg tac aaa ggt      96
235 Leu Ile Lys Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly
236           20           25           30
238 caa cca atg aca ttc aga cta tta ttg gtt gat aca cct gaa aca aag      144
239 Gln Pro Met Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys
240           35           40           45
242 cat cct aaa aaa ggt gta gag aaa tat ggt cct gaa gca agt gca ttt      192
243 His Pro Lys Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe
244           50           55           60
246 acg aaa aaa atg gta gaa aat gca aag aaa att gaa gtc gag ttt gac      240
247 Thr Lys Lys Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp
248 65           70           75           80
250 aaa ggt caa aga act gat aaa tat gga cgt ggc tta gcg tat att tat      288

```

## RAW SEQUENCE LISTING

DATE: 01/27/2006

PATENT APPLICATION: US/10/524,053A

TIME: 14:44:07

Input Set : A:\DOCSMTL-#1912723-v1-SEQLIST2.TXT

Output Set: N:\CRF4\01272006\J524053A.raw

```

251 Lys Gly Gln Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr
252                85                90                95
254 gct gat gga aaa atg gta aac gaa gct tta gtt cgt caa ggc ttg gct      336
255 Ala Asp Gly Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala
256                100                105                110
258 aaa gtt gct tat gtt tac aaa cct gaa ttc atg      369
259 Lys Val Ala Tyr Val Tyr Lys Pro Glu Phe Met
260                115                120
263 <210> SEQ ID NO: 16
264 <211> LENGTH: 123
265 <212> TYPE: PRT
266 <213> ORGANISM: Artificial Sequence
268 <220> FEATURE:
269 <223> OTHER INFORMATION: Fusion Carrier Protein named SFC120 ( N118E,
270      N119F)
272 <400> SEQUENCE: 16
273 Thr Ala Met Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr
274 1                5                10                15
275 Leu Ile Lys Ala Ile Asp Gly Asp Thr Val Lys Leu Met Tyr Lys Gly
276                20                25                30
277 Gln Pro Met Thr Phe Arg Leu Leu Leu Val Asp Thr Pro Glu Thr Lys
278                35                40                45
279 His Pro Lys Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe
280                50                55                60
281 Thr Lys Lys Met Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp
282 65                70                75                80
283 Lys Gly Gln Arg Thr Asp Lys Tyr Gly Arg Gly Leu Ala Tyr Ile Tyr
284                85                90                95
285 Ala Asp Gly Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu Ala
286                100                105                110
287 Lys Val Ala Tyr Val Tyr Lys Pro Glu Phe Met
288                115                120
290 <210> SEQ ID NO: 17
291 <211> LENGTH: 393
292 <212> TYPE: DNA
293 <213> ORGANISM: Artificial Sequence
295 <220> FEATURE:
296 <223> OTHER INFORMATION: bacterial expression vector pHSN-Met65Leu that encodes a
fusion protein, containing a sequence for a gene encoding the N-terminal nucleotide binding
domain of staph. nuclease, designated SFC120
298 <220> FEATURE:
299 <221> NAME/KEY: CDS
300 <222> LOCATION: (1)...(393)
302 <400> SEQUENCE: 17
303 aca gcc atg gga cac cat cac cat cac cat ggc gca act tca act aaa      48
304 Thr Ala Met Gly His His His His His His Gly Ala Thr Ser Thr Lys
305 1                5                10                15
307 aaa tta cat aaa gaa cct gcg act tta att aaa gcg att gat ggt gat      96
308 Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys Ala Ile Asp Gly Asp
309                20                25                30
311 acg gtt aaa tta atg tac aaa ggt caa cca atg aca ttc aga cta tta      144

```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 01/27/2006  
PATENT APPLICATION: US/10/524,053A      TIME: 14:44:08

Input Set : A:\DOCSMTL-#1912723-v1-SEQLIST2.TXT  
Output Set: N:\CRF4\01272006\J524053A.raw

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:41; Xaa Pos. 20,21,27  
Seq#:45; Xaa Pos. 22  
Seq#:48; Xaa Pos. 3,5,6,8,10,11

**Invalid Line Length:**

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:9; Line(s) 108  
Seq#:10; Line(s) 120  
Seq#:12; Line(s) 137  
Seq#:17; Line(s) 296  
Seq#:18; Line(s) 346  
Seq#:19; Line(s) 374  
Seq#:20; Line(s) 421  
Seq#:21; Line(s) 448  
Seq#:22; Line(s) 498  
Seq#:27; Line(s) 560  
Seq#:39; Line(s) 714  
Seq#:41; Line(s) 740

## VERIFICATION SUMMARY

DATE: 01/27/2006

PATENT APPLICATION: US/10/524,053A

TIME: 14:44:08

Input Set : A:\DOCSMTL-#1912723-v1-SEQLIST2.TXT

Output Set: N:\CRF4\01272006\J524053A.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application No  
L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:129 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (11) SEQUENCE:  
L:524 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (23) SEQUENCE:  
L:530 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (24) SEQUENCE:  
L:761 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:41 after pos.:16  
L:811 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:45  
L:818 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:45 after pos.:16  
L:885 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:48 after pos.:0